Elsevier Editorial System(tm) for Automation in Construction
Manuscript Draft

Manuscript Number: AUTCON-D-14-00462R4

Title: Predicting Movements of Onsite Workers and Mobile Equipment for Enhancing Construction Site Safety

Article Type: Original Paper

Keywords: Movement prediction; Kalman filtering; Construction safety

Corresponding Author: Dr. Zhenhua Zhu,

Corresponding Author's Institution:

First Author: Zhenhua Zhu

Order of Authors: Zhenhua Zhu; ManWoo Park; Christian Koch; Mohamad Soltani; Amin Hammad; Khashayar Davari

Predicting Movements of Onsite Workers and Mobile Equipment for Enhancing Construction Site Safety

Zhenhua Zhu[1], Man-Woo Park[2], Christian Koch[3], Mohamad Soltani[4], Amin Hammad[5], and Khashayar Davari[6]

Abstract:

Tens of thousands of time-loss injuries and deaths are annually reported from the construction sector, and a high percentage of them are due to the workers being struck by mobile equipment on sites. In order to address this site safety issue, it is necessary to provide proactive warning systems. One critical part in such systems is to locate the current positions of onsite workers and mobile equipment and also predict their future positions to prevent immediate collisions. This paper proposes novel Kalman filters for predicting the movements of the workers and mobile equipment on the construction sites. The filters take the positions of the equipment and workers estimated from multiple video cameras as input, and output the corresponding predictions on their future positions. Moreover, the filters could adjust their predictions based on the worker or equipment's previous movements. The effectiveness of the filters has been tested with real site videos and the results show the high prediction accuracy of the filters.

Keywords

Movement prediction; Kalman filtering; Construction safety

[1] Assistant Professor, Department of Building, Civil, and Environmental Engineering, Concordia University, Montreal, H3G 1M8, Canada, Tel: 1-514-848-2424 ext. 5948, Fax: 514-848-7965, E-mail: zhenhua.zhu@concordia.ca (*Corresponding author)
[2] Assistant Professor, Department of Civil and Environmental Engineering, Myongji University, Cheoin-gu, Yongin, Gyeonggi-do, 449-728, South Korea, Tel: 82-31-330-6411, Email: mwpark@mju.ac.kr
[3] Associate Professor, Department of Civil Engineering, the University of Nottingham, Room B27 Coates Building, University Park, Nottingham, NG7 2RD, UK, Tel: +44 (0) 115 84 68933, Email: christian.koch@nottingham.ac.uk
[4] Graduate Research Assistant, Department of Building, Civil, and Environmental Engineering, Concordia University, Montreal, H3G 1M8, Canada, Email: m_m_soltani@yahoo.com
[5] Professor, CIISE, Concordia University, Montreal, H3G 1M8, Canada, Tel: 1-514-848-2424 ext. 5800, E-mail:hammad@ciise.concordia.ca
[6] Graduate Research Assistant, Department of Building, Civil, and Environmental Engineering, Concordia University, Montreal, H3G 1M8, Canada, Email: kashdavari@gmail.com

INTRODUCTION

The construction site is typically dirty, disordered, and cluttered with different kinds of resources. Also, it is characterized by a constantly changing environment with the movement and interactions between workers and equipment. In such a chaotic and dynamic place, an incredibly high number of construction activities take place, which easily lead to construction accidents and work-related injuries and deaths. For example, in Canada, around 27,000 accepted time-loss injuries and 200 fatalities were reported in the construction sector every year from 2010 to 2012, according to the Association of Workers' Compensation Boards of Canada [1, 2]. Similarly, the U.S. Bureau of Labor Statistics noted that 183,000 construction workers were injured, and 775 workers died on the job with a fatal work injury rate of 9.5 deaths per 100,000 fulltime equivalent workers [3]. The large number of injuries and deaths makes the construction sector one of the most dangerous job sectors over the world.

Many of construction accidents are struck-by accidents, i.e. the workers being struck by mobile equipment on the construction sites [4]. The stuck-by accidents could occur, even when the workers wear high visibility clothing on the sites as required by existing safety codes and standards. In 2012, 156 fatalities due to the struck-by accidents were reported by the U.S. private construction industry [5]. In British Columbia, there were a total of 6,622 claims related to the struck-by accidents from 2006 to 2008, which represented 22% of claim volumes and 14% of claim costs resulting from construction accidents [6]. The situation becomes even worse in road construction projects, where workers might be struck by mobile equipment for construction and maintenance as well as by cars, vans, and motorcycles. 442 fatal injuries (53 percent) on road construction sites during the 2003 - 2010 periods were due to the workers being struck by vehicles or mobile equipment [3].

In order to address this site safety issue, several research studies have been proposed. They focused on the use of remote locating and tracking techniques to perform simple equipment-worker close proximity alerts. These techniques include but are not limited to Radio Frequency Identification (RFID), Ultra Wideband (UWB), Global Positioning Systems (GPS) [7]. They require remote sensors to be physically installed on the equipment and workers, so that the signals sent from the sensors could be read and interpreted. This way, the positions of the equipment and workers on the site could be located and tracked.

Compared with existing research studies, this paper relies on computer vision techniques to estimate the positions of construction workers and equipment. Moreover, the movements of the workers and equipment are predicted to get their possible positions in a short period of time. This way, the potential collisions between the workers and equipment could be avoided in a proactive way. In the paper, both position estimation and prediction parts have been integrated into one framework. Under the framework, the current positions of the equipment and workers are first estimated with the live videos collected by two or more cameras on the construction site. These positions are then input to a Kalman filter. In general, the Kalman filter is an optimal estimator that is able to infer parameters of interest from indirect, inaccurate and uncertain observations [8]. Here, the filter is specially designed to model motions (i.e. positions, velocities, and accelerations) of equipment and workers based on a series of position measurements, including noise and other inaccuracies, observed over time. The designed filter adjusts its prediction parameters with the positions newly input as well as the history of the positions estimated previously. This way, the predictions for the positions of the equipment and workers on the site could be made.

The framework in this paper does not require the installation of any remote sensors on the equipment and workers. This makes the method affordable at most construction sites, especially the large-scale ones, where hundreds of construction workers and equipment could be involved. Also, the method could be used in the case when the installation of physical sensors is not applicable. For example, in a highway construction project, the workers on the site might be struck by traffic vehicles, such as cars, vans, and motorcycles. However, it is difficult to install the physical sensors on the traffic vehicles and track their positions for the purpose of issuing the close proximity safety warnings to the workers.

The effectiveness of the proposed framework has been tested on real site videos collected by two cameras. The results showed that the average estimation errors were 0.26 meters and 0.28 meters for the movement of the worker and vehicle, while the corresponding prediction errors were 0.38 meters and 0.18 meters. The longer the predictions were made, the more accuracy the predictions could reach. The low estimation and prediction errors during the tests indicated that the proposed method in this paper could approximately estimate and predict the movement of the equipment and workers in advance. The predictions could be used to reduce the chance of struck-by accidents and therefore has the potential to enhance construction site safety. The enhancement of on-site construction safety will bring several benefits. For example, it could improve the workers' morale and job satisfactions, and increase their productivity.

Also, it could reduce project costs directly and indirectly, especially considering that the average cost per case of death or injury could reach tens of thousands of dollars in the construction industry. The prevention of one death or injury per day might lead to the cost savings of millions of dollars per year.

REMOTE LOCATING AND TRACKING FOR SITE SAFETY ENHANCEMENT

Construction researchers and safety professionals believe that existing site safety regulations are not sufficient, considering the unsatisfactory safety records in the construction industry. Therefore, it is necessary to add an extra level of safety measures to protect construction workers [9]. One of the proactive safety measures is to provide equipment-workers close proximity warnings. It means that a safety warning will be issued to an equipment operator for his/her attention, when on-foot workers are near-by [4, 10]. The close proximity warnings were expected to reduce the accidents that happened in the blind areas of equipment, as investigated by Ruff [7]. Another proactive safety measure is to create virtual fences. Typically, the virtual fences are created around known dangerous areas on the job site. If workers are approaching the areas, alarms will be issued to alert them [11 - 13].

In order to provide both proactive safety measures, it is necessary to remotely locate and track on-foot workers and mobile equipment on the construction sites. So far, several remote sensing techniques have been investigated, including GPS, RFID, UWB, etc. GPS is an outdoor satellite-based worldwide navigation system, which relies on a constellation of Earth orbiting satellites to determine the positions of GPS receivers [14]. RFID is an automatic identification technology. It is mainly used for the identification of objects on the site, but could also approximately locate them based on the radio waves communication between the RFID tags and readers [15]. UWB is a short pulse radio frequency waveform, which could provide accurate object location information based on the time-difference-of-arrival measurements [16, 17].

These remote sensing techniques mentioned above all require attaching physical signal readers and tags on the equipment and workers. For example, in the method of Marks and Teizer [4], they have an in-cab device for mobile equipment and personal device for ground workers, which contain antenna, reader, chip, battery, etc. Similarly, Ruff had the GPS antennas installed on the surface mining equipment in order to locate the equipment and evaluate its GPS-based proximity warnings [7]. If the workers and equipment need to be physically tagged, it would lead to a significant amount of additional costs for the general contractors, although the price of the tags and sensors keeps decreasing. In addition, tagging construction workers could be opposed by the unions due to the associated privacy issues and health concerns. Moreover, in a highway construction project, the workers need to be protected from traffic vehicles, such as cars, vans, and motorcycles, but it is impossible to tag, locate and track those traffic vehicles for providing the proximity warnings.

Compared with the remote sensing techniques with physical signal sensors, readers, and tags, the vision techniques could also provide the potentials to remotely locate and track the workers and equipment on the construction site. One of well-known techniques to provide three dimensional (3D) position information is referred to as stereo vision, which reconstructs the 3D position of an object through the camera calibration and triangulation principles [18]. So far, several research studies based on stereo vision have been introduced and applied in the construction field, but most of them focused on the reconstruction of static scenes. For example, Son and Kim used a stereo vision system to acquire and recognize 3D structural components [19]. Rashidi et al. relied on stereo vision to generate dense depth maps for the transportation infrastructure, such as highway bridges [20]. Fathi and Brilakis proposed a novel method for creating as-built models of sheet metal roof panels to facilitate the digital roof fabrication process with the aid of stereo vision [21].

As for enhancing site safety, Steele et al. once mount a stereo camera on the rear of an off-highway dump truck [22]. The stereo camera helped the truck driver to identify possible obstacles on the mining site [7]. Han and Lee analyzed workers' unsafe actions that may cause incidents (e.g. fall from a ladder due to leaning too far to one side or reaching too far overhead) from the videos captured by stereo cameras [23]. Weerasinghe and Ruwanpura developed a conceptual model, Automated Multiple Objects Tracking System, to track construction objects, such as workers and tools, with fixed video surveillance cameras [24].

One main benefit of using vision techniques to locate and track construction workers and equipment is that the workers and equipment do not have to be physically tagged. Therefore, several issues related to physically tagging the workers and equipment in the remote sensing techniques could be addressed. Also, it becomes more and more common to place the cameras around the site to capture job site activities and record project construction progress [25]. The cameras could take pictures or videos with a high resolution and wide field of view. Therefore, the workers equipment, and even non project-related entities, such as traffic vehicles in highway construction projects, could be remotely monitored with a limited number of cameras.

OBJECTIVE AND SCOPE

The ultimate goal of this ongoing research work is to investigate the feasibility of creating a proactive, real-time safety alert system with the live video frames from construction cameras. In order to achieve this goal, it is necessary to estimate the current 3D positions of the workers and equipment. Also, it is important to predict their future movements. Consider the recent writers' work on estimating 3D positions of the workers and equipment [26], which will be briefly described later. The specific focus of this paper is placed on evaluating whether their future positions could be reasonably predicted based on their previous estimated positions. If the tests show the prediction results are also promising, both positions estimation and prediction together will build a solid foundation for creating a vision-based proactive, real-time safety alert system to provide equipment-workers close proximity warnings and creating virtual fences on the construction sites.

The work presented in this paper does not intend to enhance the visibility of onsite construction cameras. It is assumed to function when the videos collected by the cameras are clear with acceptable quality and a limited degree of occlusions. The occlusions could be one of the major obstacles that affect the performance of vision techniques. However, this issue could be addressed or at least alleviated by installing the cameras at a certain level of height and carefully selecting the camera placements on the construction sites.

In addition, this research work does not plan to replace the role of onsite inspectors, such as construction site health and safety management guarantors in Quebec. Those inspectors are responsible to identify and address potential onsite safety issues, if there are any. Therefore, this research work is not to replace them but facilitate their onsite work by helping them monitor construction workers and equipment, and predict their motions with real-time feedbacks.

PROPOSED FRAMEWORK

In order to achieve the above-mentioned objective, a novel vision-based framework has been proposed here. The framework includes two main steps, as illustrated in Figure 1. Under the framework, two or more construction cameras are placed at a construction site to monitor job site activities from different angles. The site videos captured by the cameras are transferred to a workstation for analysis. There, the onsite positions of the workers and equipment in the videos are estimated using the triangulation principle. Based on the estimated positions, the future positions of the workers and equipment on the site are predicted through the Kalman filtering [27]. Moreover, the prediction parameters in the Kalman filter are frequently updated by comparing its predictions with the onsite positions estimated later.

<Insert Figure 1 here>

Positions Estimation from Multi-View Videos

The estimation of the 3D positions from videos mainly follows the procedure proposed by Park et al. [26], which includes 1) camera calibration, 2) pose estimation, 3) visual detection and tracking and 4) triangulation (Figure 2). Both camera calibration and pose estimation are performed offline, while the work of visual detection and tracking and triangulation are done online. When the cameras are installed on the construction site, it is necessary to make sure they have partially overlapping views of the site. The cameras are then calibrated using Bouguet's calibration toolbox [28] to calculate their intrinsic parameters (focal length, lens distortion, etc.). Also, the external orientation and position of one camera in relation to another are estimated and represented as a rotation matrix (R) plus a translation vector (t). Moreover, the essential matrix is computed using the normalized eight-point algorithm [18]. The points required in the algorithm are extracted and matched with the Scale-Invariant Feature Transform (SIFT) [29] combined with the Maximum a Posteriori Sample Consensus (MAPSAC) [30] to remove potential feature outliers.

<Insert Figure 2 here>

After the camera calibration and pose estimation, the 3D positions of the equipment and workers on the construction site could be automatically estimated through visual detection, tracking, and triangulation. First, the construction workers and equipment are detected based on their respective visual features. The detection results then initialize a kernel-based 2D tracking algorithm [31], which could track the detected workers and equipment subsequently in each site video frame. The video-based tracking results produce 2D centroids in each video frame, which indicate the positions of the workers and equipment in the videos. The 2D centroids are combined with the camera intrinsic and extrinsic parameters through the triangulation. This way, the 3D positions of the workers and equipment on the construction site could be estimated.

Positions Prediction through the Kalman Filtering

The measured 3D positions are fed into a Kalman filter to predict the positions of the workers and equipment at the next moment. In order to prepare the filter, first, the state of the worker or equipment at time step t is expressed as a vector (Eq. 1), which includes the positions (x, y, z), velocities $(\dot{x}, \dot{y}, \dot{z})$, and accelerations $(\ddot{x}, \ddot{y}, \ddot{z})$ along the three coordinate axes. Then, the dynamics of the worker's or equipment's motion on the construction site is modeled as a time-invariant system (Eq. 2)

$$S_t = (x_t, y_t, z_t, \dot{x}_t, \dot{y}_t, \dot{z}_t, \ddot{x}_t, \ddot{y}_t, \ddot{z}_t)^T \tag{1}$$

$$\frac{dS_t}{dt} = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \times S_t + \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \times W_t \tag{2}$$

where S_t is the system state at step t and W_t is a white noise process with power spectral density. Suppose Δt is the time step size of two consecutive measurements. This way, the state transition matrix A_t could be defined in Eq. 3. Meanwhile, the measurement matrix is correspondingly set (Eq. 4), since the only measurement available is the 3D positions of the worker or equipment without any information of the velocities and accelerations.

$$A = \begin{pmatrix} 1 & 0 & 0 & \Delta t & 0 & 0 & 0.5\Delta t^2 & 0 & 0 \\ 0 & 1 & 0 & 0 & \Delta t & 0 & 0 & 0.5\Delta t^2 & 0 \\ 0 & 0 & 1 & 0 & 0 & \Delta t & 0 & 0 & 0.5\Delta t^2 \\ 0 & 0 & 0 & 1 & 0 & 0 & \Delta t & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & \Delta t & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & \Delta t \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{3}$$

$$H = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \tag{4}$$

After the preparation of the filter, the next state of the system is predicted by the filter based on the previous measurements and the state transition matrix. Also, the prediction results are compared with the real measurements at the next moment. The difference between the two is further used to update the filter for the sake of correcting its predictions in the future. The prediction and update processes could be described with the following equations (Eq. 5 - 10) (Welch and Bishop, 1997).

- *Prediction:*

$$S_t^- = A \times S_{t-1} \tag{5}$$

$$P_t^- = A \times P_{t-1} \times A^T + Q \tag{6}$$

- *Update::*

$$K_t = P_t^- \times H^T \times \left(H \times P_t^- \times H^T + R\right)^{-1} \tag{7}$$

$$v_t = y_t - H \times S_t^- \tag{8}$$

$$S_t = S_t^- + K_t \times v_t \tag{9}$$

$$P_t = P_t^- - K_t \times \left(H \times P_t^- \times H^T + R\right) \times K_t^T \tag{10}$$

where S_t^- and S_t are the predicted and estimated mean of system states before and after seeing the real measurements; P_t^- and P_t are the predicted and estimated covariance of the system states before and after seeing the real measurements; Q is the process noise covariance; R is the measurement noise covariance; v_t is the measurement residual on time step t; K_t is defined as the filter gain, which indicates how much corrections should be made on time step t.

Figure 3 illustrates the overall process for the position prediction and update with the Kalman filtering. Specifically, Eq. 5 and 6 are predictor equations. They are used to compute the predicted mean and error covariance of the motion system to obtain the priori position estimates for the next time step. Eq. 7 – 10 are corrector equations. They are responsible for obtaining a posteriori position estimate, when the new position measurement is incorporated. The first step during the update is to compute the Kalman gain (Eq. 7). Then, the actual measurement of the motion system is made and incorporated to generate a posteriori estimate for the system (Eq. 8 and 9). The final step is to estimate a posteriori error covariance (Eq. 10). The process for the prediction and update is repeated with the previous posterior estimates to predict the new priori estimates in a recursive nature.

<Insert Figure 3 here>

EXPERIMENTS AND RESULTS

The methods in the proposed framework were tested with the videos recorded by two high-definition (HD) camcorders, Canon VISXIAHF S100, under the resolution of 1,920 × 1,080 pixels at 30 frames per second. The camcorders were located to collect the video frames of the construction site, where a facility was to be built for indoor football practices. The site was managed by Barton Malow Company. In order to get the stereo videos, the cameras were placed separately at the distance of 8.3 meters apart from each other. The relative positions between the cameras, worker, and vehicle have been illustrated in Figure 4.

<Insert Figure 4 here>

Figure 5 shows the examples of the video frames collected by the two cameras. These video frames recorded the movement of a worker and a vehicle on the construction site. Based on the video frames, the 3D positions of the worker and vehicle were estimated. These positions were compared with the position information collected by a total station to determine the estimation accuracy. The overall effectiveness of estimating the 3D positions of the worker and vehicle has been summarized in Table 1. It was found the average errors of estimating the 3D positions of a worker and vehicle from two video cameras were 0.26 meters and 0.28 meters with the standard deviations of 0.19 meters and 0.19 meters respectively. The maximum estimation errors were limited to 1.05 meters for a worker and 0.90 meters for a vehicle. More details about the experiments and results could be found in the recent work of Park et al. [26].

<Insert Figure 5 here>

<Insert Table 1 >

The positions prediction work took the 3D positions measured from the videos before as input and produced the predictions at each time step as output. Figure 6 and 7 compared the 3D positions measurements and predictions for the movement of the construction worker and vehicle in 2D views (X-Z plane). The numerical comparison results have been summarized in Table 2 and 3. Compared with the measurements, it was found that the mean error in predicting the movement of the worker was 0.32 meters with the standard deviation of 2.38 meters, and the mean error in predicting the movement of the vehicle was 0.18 meters with the standard deviation of 1.08 meters. More specifically, the mean errors in X-, Y-, and Z- directions were 0.06 meters, 0.08 meters, and 0.28 meters with the standard deviations of 0.03 meters, 0.70 meters, and 2.28 meters, when predicting the worker's movement. The mean errors in X-, Y-, and Z- directions were 0.06 meters, 0.08 meters, and 0.28 meters with the standard deviations of 0.04 meters, 0.02 meters, and 0.16 meters, when predicting the vehicle's movement.

<Insert Figure 6 here>

<Insert Figure 7 here>

<Insert Table 2 here>

<Insert Table 3 here>

The large prediction errors were typically made at the initial prediction stage. For example, it was noted in Table 2 that the maximum error from the first 90 predictions was 55.91 meters, and the maximum prediction errors in Y-, and Z- directions could reach 16.36 and 53.44 meters, when predicting the worker's movement. Similarly, when predicting the vehicle's movement, the maximum errors from the first 90 predictions was 32.61 meters, and the maximum prediction errors in Y-, and Z- directions could reach 2.88 and 32.48 meters. This is mainly because the designed Kalman filter did not have the sufficient "prior knowledge" of the movement of the workers and/or equipment to make accurate predictions.

The "prior knowledge" could be automatically accumulated by the filter. During the tests, the filter updated its parameters through identifying and correcting its previous prediction mistakes. This way, the knowledge to make accurate predictions was learned. Typically, the learning process was done in a fast way. Consider the cameras captured 30 video frames per second (FPS). It means that it was possible to make 30 measurements in one second. Therefore, the initial 90 predictions could be done to cover the movement of the worker or vehicle in their initial 3 seconds.

When the sufficient "prior knowledge" has been obtained, the predictions made by the filter reached a reasonable accuracy. As illustrated in Table 2, the maximum prediction errors in X-, Y- and Z- directions were limited to 0.15 meters, 0.05 meters, and 0.20 meters for predicting the worker's movement, if the first 90 predictions were ignored. Correspondingly, the maximum error in 3D was reduced to be 0.22 meters. As for predicting the vehicle's movement, the maximum errors of the movement perditions in X-, Y- and Z- directions were limited to 0.25 meters, 0.09 meters, and 0.56 meters, and the maximum error in 3D was 0.56 meters (Table 3).

As illustrated in Figure 4, the cameras were set up about 30 ~ 40 meters away from the worker. When the measurements and predictions are made at 30 frames per second (fps) by default, the prediction error could reach 0.02 meters after initial 90 predictions. The prediction error is increased with the reduction of the frequency for the measurements and predictions. Figure 8 showed that the errors for predicting worker’s movement would increase to 0.44 meters, 0.73 meters, and 1.58 meters, when the measurements and predictions are made every 0.5, 1, and 1.5 seconds. Similar findings were also noted when predicting the movement of the vehicle in the tests.

<Insert Figure 8 here>

CONCLUSIONS AND FUTURE WORK

This paper designed Kalman filters to predict the future positions of onsite workers and mobile equipment. The predictions were made based on the current positions of the workers and equipment on the sites and also their previous movement records. The prediction results could indicate the movements of the workers and equipment in a short period of time from the current moment. This information is useful to create a proactive warning system to prevent immediate potential collisions on the construction site and therefore enhance construction site safety.

The Kalman filters designed in the paper has been tested with real site videos. The test results showed that the position predictions made by the filters could reflect the real movement of the worker and equipment. Specifically, the average errors in predicting the worker's and vehicle's movements could reach 0.38 meters and 0.18 meters. More accurate predictions could be achieved, when the Kalman filter got sufficient knowledge from its previous prediction errors. For example, the average prediction errors for the worker's and vehicle's movements could be reduced to 0.10 meters and 0.11 meters, when the first 90 predictions within approximately 3 seconds were ignored. The high prediction accuracy indicated the effectiveness of the Kalman filters designed in this paper. Future work will focus on creating a pro-active collision warning system based on the work presented in this paper.

Future work will be focused on two aspects. First, more experiments will be performed to test the tolerance of the predictions made by the work in this paper on various motion routes. Also, a pro-active collision warning system will be developed at construction jobsites to check the cost effectiveness of implementing the system in construction projects. The authors have been working with the local industry to create a multi-camera environment on a construction site in Montreal. The site will be used as a test bed to implement the collision warning system. Compared with existing safety enhancement research studies with the reliance on remote sensing techniques, the system relies on the videos remotely captured by high-definition cameras. It is not necessary to physically install or put any sensors or tags on the workers and equipment, which is supposed to make the system more affordable.

REFERENCES

[1] Association of Workers' Compensation Boards of Canada (AWCBC) (2012a). "Number of Accepted Time-Loss Injuries, by Industry and Jurisdiction, 2010-2012." <http://awcbc.org/?page_id=14#injuries> (March 29, 2014)

[2] Association of Workers' Compensation Boards of Canada (AWCBC) (2012b). "Number of Fatalities, by Industry and Jurisdiction, 2010-2012." <http://awcbc.org/?page_id=14#fatalities> (March 29, 2014)

[3] Bureau of Labor Statistics (BLS) (2013). "Monthly Labor Review - An Analysis of Fatal Occupational Injuries at Road Construction Sites, 2003-2010." <http://www.bls.gov/opub/mlr/2013/article/an-analysis-of-fatal-occupational-injuries-at-road-construction-sites-2003-2010.htm> (March 30, 2014)

[4] Marks, E. and Teizer, J. (2013). "Method for testing proximity detection and alert technology for safe construction equipment operation." Construction Management and Economics, 31(6): 636-646.

[5] Bureau of Labor Statistics (BLS) (2012). "2012 Census of Fatal Occupational Injuries: Fatal occupational injuries by industry and selected event or exposure." Bureau of Labor Statistics, < http://www.bls.gov/iif/oshcfoi1.htm> (Sept. 25, 2014).

[6] WorkSafeBC (2014). "Statistics for Construction - Claims by Accident Type, 2006 - 2008." Safety at Work, WorkSafeBC, <http://www2.worksafebc.com/Portals/Construction/Statistics.asp> (Sept. 25, 2014)

[7] Ruff, T. (2007). "Recommendations for evaluating and implementing proximity warning systems on surface mining equipment, Report of Investigations 9672, National Institute for Occupational Safety and Health.

[8] Lindsey, T. (2014). “On the Kalman Filter and its Applications.” Master Thesis, Department of Mathematics, University of Kansas, Lawrence, USA, <https://kuscholarworks.ku.edu/bitstream/handle/1808/14535/Lindsey_ku_0099M_13455_DATA_1.pdf?sequence=1> (Jan. 13, 2016).

[9] Green, L. and Tominack, G. (2012). "Real-time proactive safety in construction." POWER Magazine, <http://www.powermag.com/real-time-proactive-safety-in-construction/> (August 3, 2013)

[10] Teizer, J., Allread, B.S., Fullerton, C.E. and Hinze, J. (2010) Autonomous pro-active real-time construction worker and equipment operator proximity safety alert system. Automation in Construction, 19(5), 630–40.

[11] Carbonari, A. and Giretti, A. and Naticchia, B. (2011). "A proactive system for real-time safety management in construction sites." Automation in Construction, 20(6): 686-698.

[12] Giretti, A., Carbonari, A, Naticchia, B. and Grassi, M.D. (2009). "Design and first development of an automated real-time safety management system for construction sites." Journal of Civil Engineering and Management, 15(4), 325-336. DOI:10.3846/1392-3730.2009.15.325-336

[13] Zhang, C., Hammad, A., Soltani, M., Setayeshgar, S., and Motamedi, A. (2012). "Dynamic virtual fences for improving workers safety using BIM and RTLS." In:

Proc. of 14th International Conference on Computing in Civil and Building Engineering, Moscow, Russia, 24-29 June 2012.

[14] Drira, A. (2006). "GPS Navigation for Outdoor and Indoor Environments." M.A.Sc. Thesis, University of Tennessee, Knoxville, USA.

[15] Angell I. and Kietzmann, J. (2006). "RFID and the end of CASH?" Communications of the ACM, 49(12), 90-96.

[16] McKinney, J. Lin, I., Weiner, A. (2006). "Shaping the Power Spectrum of Ultra-Wideband Radio-Frequency Signals." IEEE Transactions on Microwave Theory and Techniques, 54(12), 4247-4255.

[17] Cheng, T., and Teizer, J. (2013). "Real-time resource location data collection and visualization technology for construction safety and activity monitoring applications." Automation in Construction, 34, 3-15.

[18] Hartley, R., and Zisserman, A. (2004). Multiple view geometry in computer vision, Cambridge University Press, Cambridge, UK.

[19] Son, H., and Kim, C. (2010). “3D structural component recognition and modeling method using color and 3D data for construction progress monitoring.” Automation in Construction, 19(7), 844–854.

[20] Rashidi, A., Fathi, H. and Brilakis, I. (2011) "Innovative Stereo Vision-Based Approach to Generate Dense Depth Map of Transportation Infrastructure" Transportation Research Record: Journal of the Transportation Research Board, Transportation Research Board of the National Academies, Volume 2215, Pages 93 – 99, DOI: 10.3141/2215-10

[21] Fathi, H. and Brilakis, I. (2013). "A videogrammetric as-built data collection method for digital fabrication of sheet metal roof panels." Advanced Engineering Informatics, 27(4), 466–476

[22] Steele, J., Derunner, C., and Whitehorn, M., (2003). “Stereo images for object detection in surface mine safety applications.” Western Mining Resource Center Tech Report No. TR20030109.

[23] Han, SU., and Lee, SH. (2013) "A vision-based motion capture and recognition framework for behavior-based safety management." Automation in Construction, 35, 131-141.

[24] Weerasinghe, I. and Ruwanpura, J. (2010). "Automated Multiple Objects Tracking System (AMOTS)." In: Proc. of Construction Research Congress 2010, Banff, Alberta, Canada, 8-10 May 2010, pp. 11-20. doi: 10.1061/41109(373)2

[25] Bohn, J. and Teizer, J. (2010). ”Benefits and Barriers of Construction Project Monitoring Using High-Resolution Automated Cameras.” J. Constr. Eng. Manage., 136(6), 632–640.

[26] Park, M.-W., Koch, C., and Brilakis, I. (2012) “Three-Dimensional Tracking of Construction Resources Using an On-Site Camera System.” Journal of Computing in Civil Engineering, 26 (4), 541-549.

[27] Welch, G. and Bishop, G. (1997). "An Introduction to the Kalman Filter." TR 95-041, Department of Computer Science, University of North Carolina at Chapel Hill, Chapel Hill, USA, <http://www.cs.unc.edu/~welch/media/pdf/kalman_intro.pdf> (Sept 25, 2014).

[28] Bouguet, J. Y. (2004). "Camera calibration toolbox for Matlab." <http://www.vision.caltech.edu/bouguetj/calib_doc> (Apr. 18, 2011).

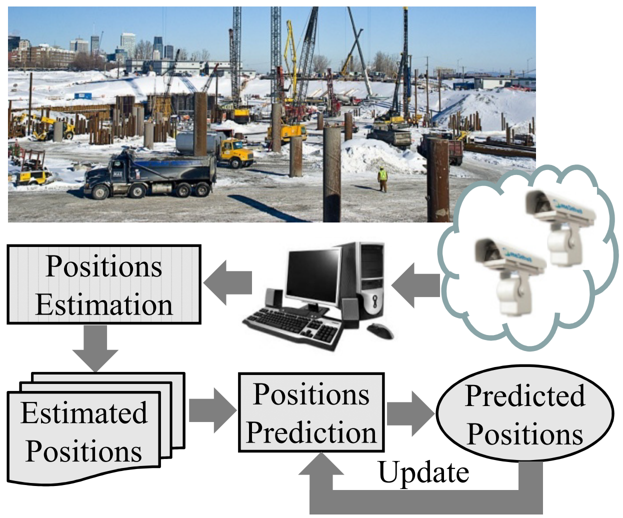
[29] Lowe, D. G. (2004). "Distinctive image features from scale-invariant keypoints." Int. J. Comput. Vis., 60(2), 91–110.

[30] Torr, P. H. S. (2002). "Bayesian model estimation and selection for epipolar geometry and generic manifold fitting." Int. J. Comput. Vis., 50(1), 35–61.

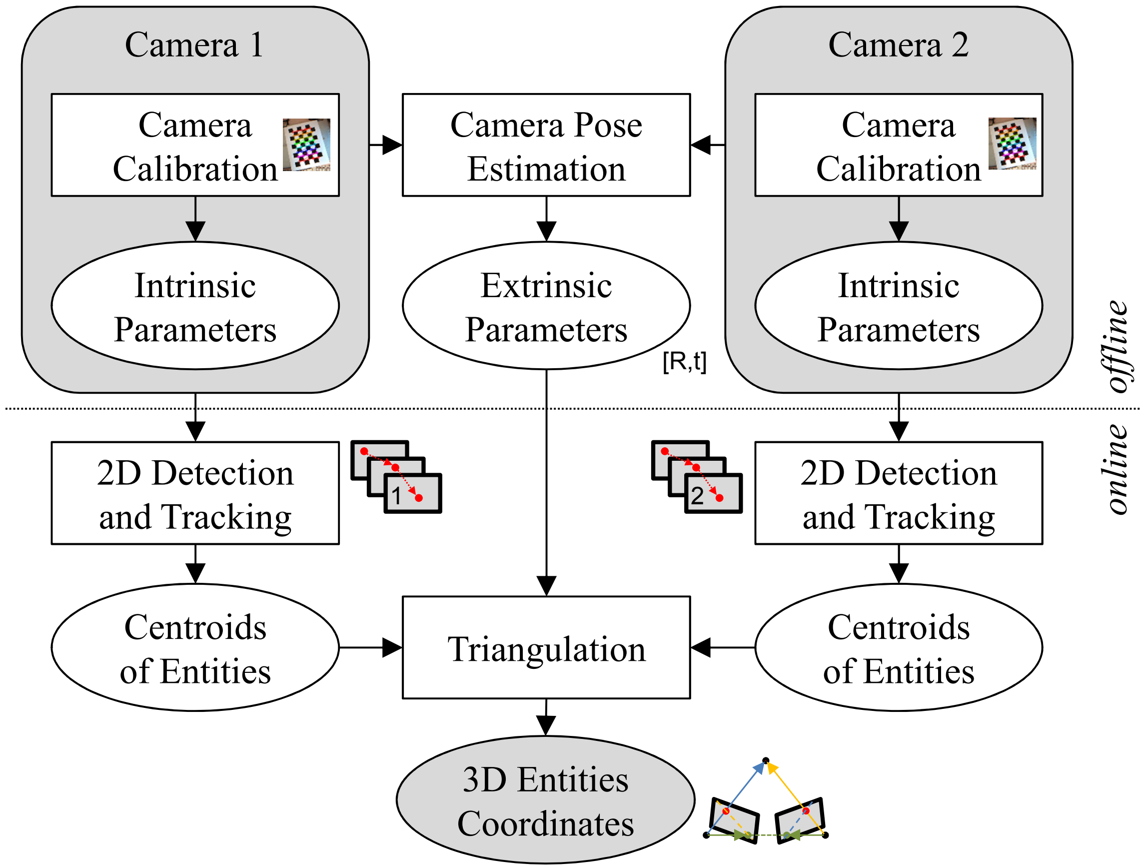
[31] Ross, D., Lim, J., Lin, R.-S., and Yang, M.-H. (2008). "Incremental learning for robust visual tracking." Int. J. Comput. Vis., 77(1), 125–141.

LIST OF FIGURES

Figure

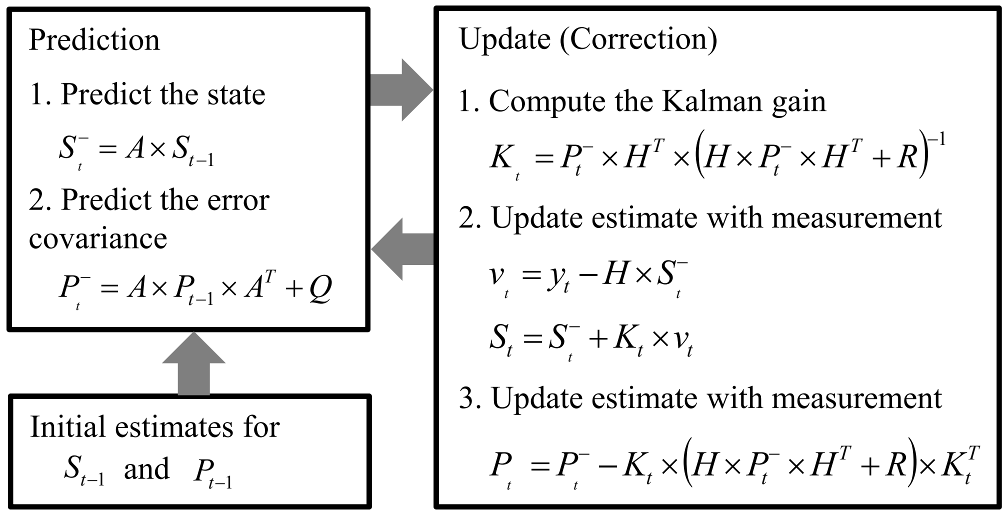

Positions Estimation
Estimated Positions
Positions Prediction
Predicted Positions
Update

Figure



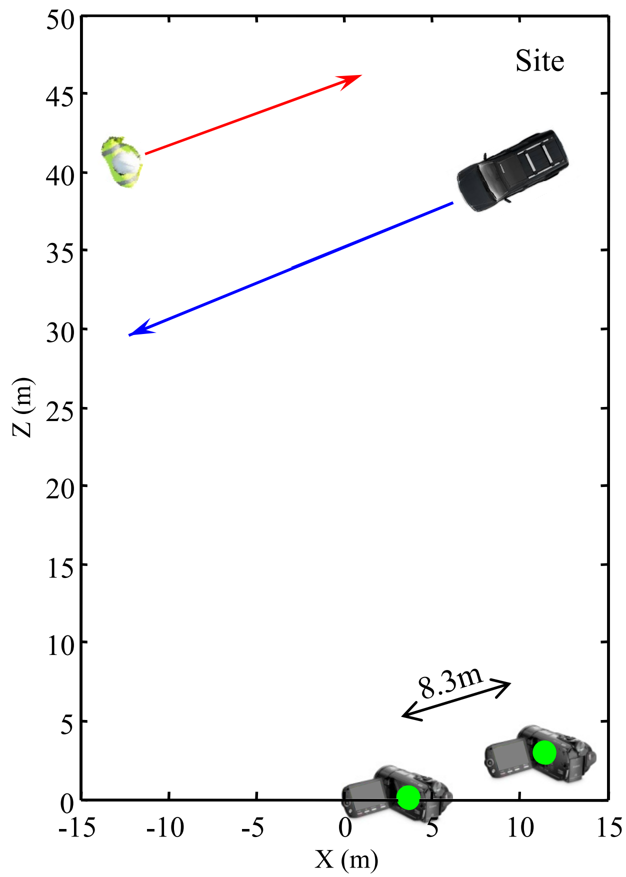
Camera 1
Camera Calibration
Intrinsic Parameters
Camera Pose Estimation
Extrinsic Parameters
[R,t]
Camera 2
Camera Calibration
Intrinsic Parameters
offline
online
2D Detection and Tracking
1
2
2D Detection and Tracking
Centroids of Entities
Triangulation
Centroids of Entities
3D Entities Coordinates

Figure

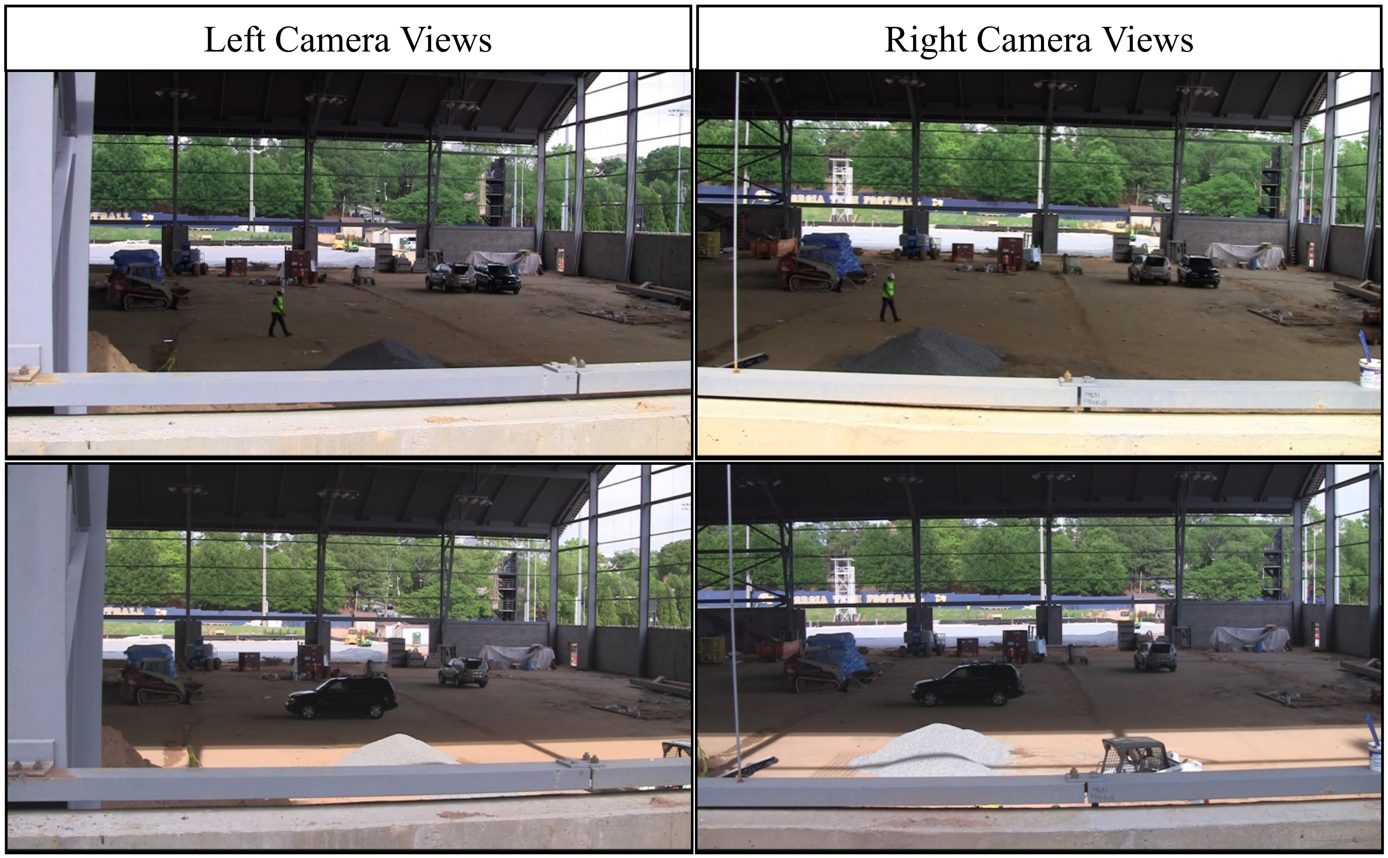


Prediction
1. Predict the state
$S_t^- = A \times S_{t-1}$
2. Predict the error covariance
$P_t^- = A \times P_{t-1} \times A^T + Q$
Initial estimates for
S_{t-1} and P_{t-1}
Update (Correction)
1. Compute the Kalman gain
$K_t = P_t^- \times H^T \times \left(H \times P_t^- \times H^T + R\right)^{-1}$
2. Update estimate with measurement
$v_t = y_t - H \times S_t^-$
$S_t = S_t^- + K_t \times v_t$
3. Update estimate with measurement
$P_t = P_t^- - K_t \times \left(H \times P_t^- \times H^T + R\right) \times K_t^T$

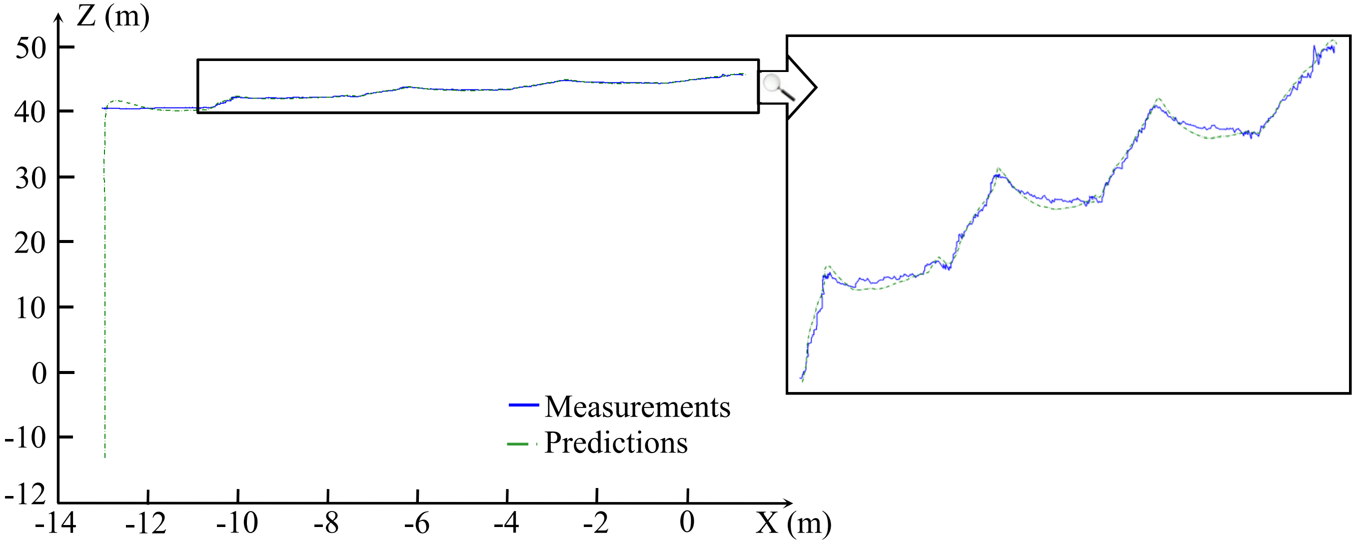
Figure



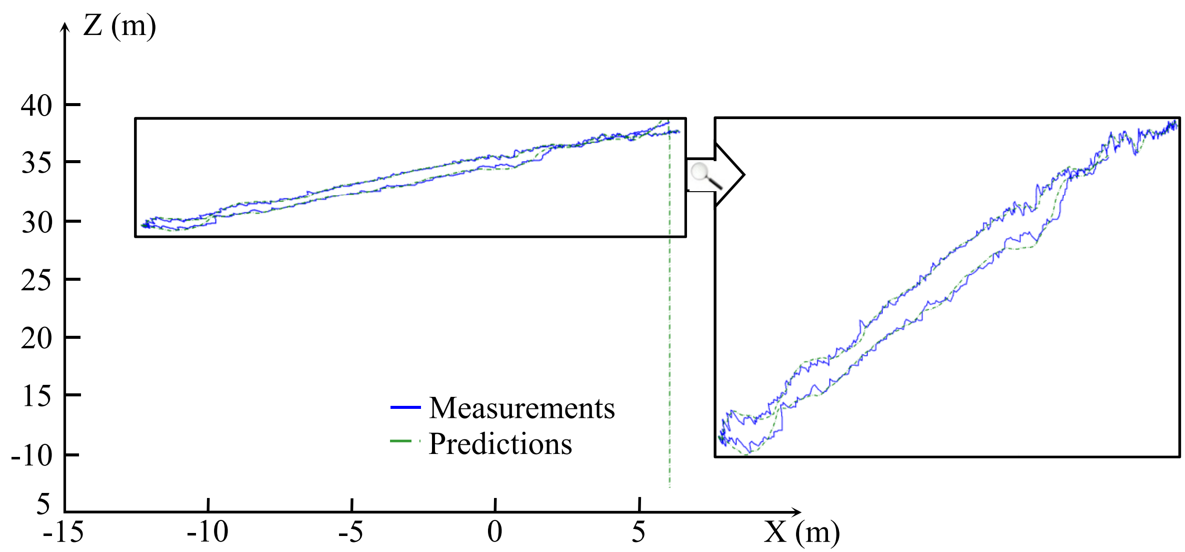
Figure



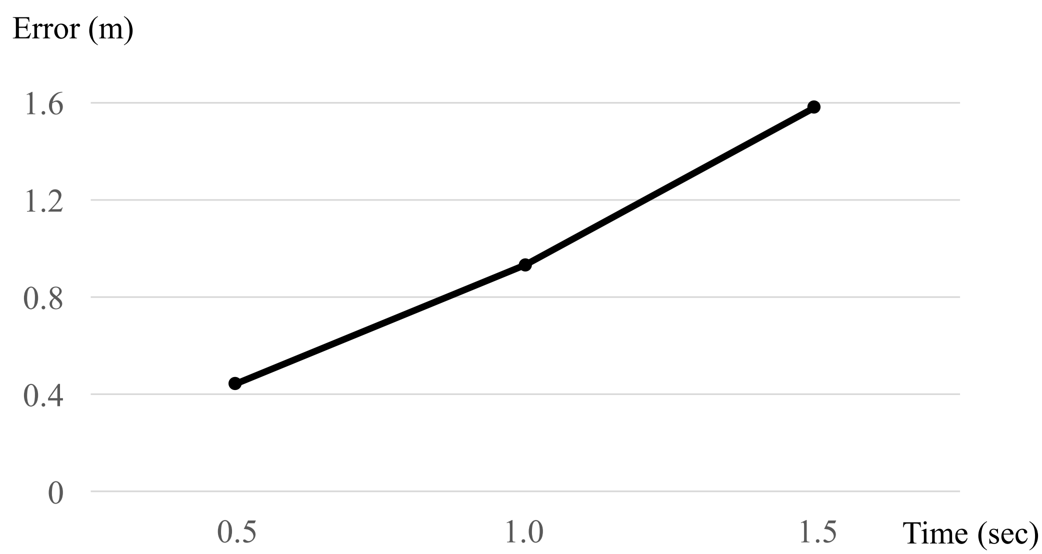
Figure


Z (m)
50
40
30
20
10
0
-10
-12
-14
-12
-10
-8
-6
-4
-2
0
X (m)
Measurements
Predictions

Figure



Figure



LIST OF TABLES

Table 1: Errors of Estimating 3D Positions using Stereo Vision System

Object Type	Error (m)		
	Max	Mean	Std.
Worker	1.05	0.26	0.19
Vehicle	0.90	0.28	0.19

Table 2: Errors in Predicted 3D Positions in Worker Movement

Errors (m)	Initial 90 predictions				Remaining predictions				All	
	Max	Min	Mean	Std.	Max	Min	Mean	Std.	Mean	Std.
X-Direction	0.16	0.00	0.06	0.04	0.15	0.00	0.06	0.03	0.06	0.03
Y-Direction	16.36	0.01	0.46	1.80	0.05	0.00	0.01	0.01	0.08	0.70
Z-Direction	53.46	0.01	1.50	5.88	0.20	0.00	0.08	0.05	0.28	2.28
3D Distance	55.91	0.02	1.58	6.14	0.22	0.02	0.10	0.04	0.32	2.38

Table 3: Errors in Predicted 3D Positions in Vehicle Movement

Errors (m)	Initial 90 predictions				Remaining predictions				All	
	Max	Min	Mean	Std.	Max	Min	Mean	Std.	Mean	Std.
X-Direction	0.06	0.00	0.02	0.02	0.25	0.00	0.05	0.04	0.04	0.04
Y-Direction	2.88	0.00	0.08	0.32	0.09	0.00	0.01	0.01	0.02	0.10
Z-Direction	32.48	0.00	0.93	3.57	0.56	0.00	0.09	0.08	0.16	1.08
3D Distance	32.61	0.02	0.93	3.58	0.56	0.01	0.11	0.09	0.18	1.08